Structural features specific to plant metallothioneins

Freisinger, Eva

Abstract: The metallothionein (MT) superfamily combines a large variety of small cysteine-rich proteins from nearly all phyla of life that have the ability to coordinate various transition metal ions, including Zn(II), Cd(II), and Cu(I). The members of the plant MT family are characterized by great sequence diversity, requiring further subdivision into four subfamilies. Very peculiar and not well understood is the presence of rather long cysteine-free amino acid linkers between the cysteine-rich regions. In light of the distinct differences in sequence to MTs from other families, it seems obvious to assume that these differences will also be manifested on the structural level. This was already impressively demonstrated with the elucidation of the three-dimensional structure of the wheat E(c)-1 MT, which revealed two metal cluster arrangements previously unprecedented for any MT. However, as this structure is so far the only one available for the plant MT family, other sources of information are in high demand. In this review the focus is thus set on any structural features known, deduced, or assumed for the plant MT proteins. This includes the determination of secondary structural elements by circular dichroism, IR, and Raman spectroscopy, the analysis of the influence of the long linker regions, and the evaluation of the spatial arrangement of the sequence separated cysteine-rich regions with the aid of, e.g., limited proteolytic digestion. In addition, special attention is paid to the contents of divalent metal ions as the metal ion to cysteine ratios are important for predicting and understanding possible metal-thiolate cluster structures.

DOI: https://doi.org/10.1007/s00775-011-0801-z

Structural features specific to plant metallothioneins

Eva Freisinger

Eva Freisinger (✉)

Institute of Inorganic Chemistry, University of Zurich,

8057 Zurich, Switzerland

e-mail: freisinger@aci.uzh.ch

Abstract The metallothionein (MT) superfamily combines a large variety of small cysteine-rich proteins from nearly all phyla of life that have the ability to coordinate various transition metal ions including Zn^{II}, Cd^{II}, and Cu^{I}. The members of the plant MT family are characterized by great sequence diversity, requiring further subdivision into four subfamilies. Very peculiar and not well understood is the presence of rather long Cys-free amino acid linkers between the Cys-rich regions. In light of the distinct differences in sequence to MTs from other families it seems obvious to assume that these differences will be also manifested on the structural level. This was already impressively demonstrated with the elucidation of the three-dimensional structure of the wheat E_c-1 MT, which revealed two metal cluster arrangements previously unprecedented for any MT. However, as this structure is so far the only one available for the plant MT family other sources of information are of high demand. In this review the focus is thus set on any structural features known, deduced or assumed for the plant MT proteins. This includes the determination of secondary structural elements by CD, IR, and Raman spectroscopy, the analysis of the influence of the long linker regions, as well as the evaluation of the spatial arrangement of the sequence separated Cys-rich regions with the aid of e.g. limited proteolytic digestion. In addition, special attention is paid to the contents of divalent metal ions as the metal ion-to-Cys ratios are important for predicting and understanding possible metal-thiolate cluster structures.

Keywords Plant metallothioneins · Metal-thiolate cluster · Linker regions · Secondary structural elements · Spatial domain arrangement

Introduction

The first report about a mammalian metallothionein (MT) dates back more than 50 years [1]. It was the high cadmium content of horse kidney cortex that caught the researchers' attention and led to the identification of the first "naturally" Cd^{II}-binding protein. Owing to its additional richness in Cys-thiolates, the name metallothionein was born. Considerably younger are the plant MTs, which were discovered around 25 years ago [2]. Here it was the observation that during the earliest germination stage of *Triticum aestivum* (common bread wheat) 20–25 % of the total Cys incorporated into nascent proteins is found in just a single protein, which was accordingly denoted “early cysteine-labeled” or E_c protein. In the following years, this peculiar wheat E_c-1 protein was investigated in more detail [3, 4], and beginning of the 1990's the overexpression of another plant MT, the MT2 protein from *Pisum sativum* (garden pea), was attempted in *Escherichia coli* but hampered by partial proteolytic digestion [5]. In the following ten years more and more plant MT gene sequences were reported and gene expression studies were performed, however, no investigations on the actual proteins were performed. This only started to change in the early 2000's, when the use of larger purification tags such as glutathione-S-transferase (GST) became more common and proteolytic digestion of especially the linker regions of these rather small proteins in *E. coli* could be prevented [6, 7]. It is thus appropriate to say that the study of plant MT proteins is still a quite recent research field. But what discriminates the plant MT family from the mammalian forms? The original criteria used to designate Cys- and metal ion rich proteins as metallothioneins were: A low molecular weight (M_r ~5−10 kDa), no aromatic amino acids, a characteristic Cys distribution pattern, e.g. CysCys and CysXaaCys motifs with Xaa marking any amino acid other than Cys, and the characteristic optical features of metal-thiolate complexes, e.g. a shoulder of the peptide backbone transitions around 250 nm indicative for Cd^{II}-thiolate clusters [8, 9]. These criteria were only met by the E_c proteins, provided that the

two highly conserved His residues found in their sequences were disregarded (Fig. 1).

>> insert Fig. 1 here (double column) <<

For the plant MT1, MT2, and MT3 proteins the presence of the aromatic amino acids Tyr and Phe caused their classification as MT-like. Only upon introduction of the presently used classification system in 1999 all plant MT forms were combined in the MT family 15, the plant MTs [10]. While the mammalian MTs invariably contain 20 Cys residues, the Cys content in plant MTs is lower and shows a greater variety across the different subfamilies. The E_c or MT4 proteins have with 17 Cys residues the highest content of thiolate groups, followed by the MT2 forms with 14, the MT1 forms with 12, and finally the MT3 proteins with only 10 Cys residues (Fig. 1). The Cys content within a plant MT subfamily can vary slightly and some such variants are depicted in Fig. 1 and marked with an asterisk. However, the length of the amino acid sequences and hence the mass of the proteins is generally higher than observed for the mammalian MTs. While the latter have masses around 6.0 kDa without metal ions, the largest plant MTs are found in the MT2 subfamily (~7.9 kDa), followed by the E_c (~7.7 kDa) and MT1 (~7.6 kDa) forms, and finally the MT3 proteins (~6.8 kDa). The observed mass differences mainly originate from the so-called linker regions, amino acid stretches connecting the Cys-rich regions of the proteins but being devoid of Cys residues themselves. In the mammalian forms these linker regions comprise only three amino acids, but in the plant MT forms they are considerably longer (Fig. 1). It is within these linker regions where the aromatic amino acids so untypical for the "classical" MTs are located. The plant MT1, MT2, and MT3 forms feature two Cys-rich regions. The six Cys residues located at the C-terminus in each subfamily are arranged in a highly conserved pattern (CxCxxxCxCxxCxC). In contrast the number and distribution pattern of the Cys residues in the N-terminal part are characteristic for each subfamily and used as the main distinguishing factor for subfamily discrimination. The E_c proteins stick out as they contain three Cys-rich regions, all clearly separated in sequence by linker regions. These linker regions are

considerably shorter than the ones found in the plant MT1, MT2, and MT3 proteins, but with 12–15 amino acids nevertheless distinctively longer than in the mammalian forms. Noticeable are also the two fully conserved His residues in the E_c subfamily mentioned already above, which are located close to and within the central Cys-rich region.

The current state of research with respect to gene expression pattern and proposed functions of plant MTs has been summarized recently [11, 12] and will not be part of the current review. Instead, after giving an update of the metal ion contents observed in the few plant MTs studied in form of the translated proteins so far, the main focus will be set on the structural aspects that can be derived from sequence analyses and spectroscopic measurements, and also aspects concerning the structural and functional role of the peculiar linker sequences typical for the plant MT forms will be addressed.

Metal ion contents

The divalent metal ions Zn^{II} and Cd^{II} are bound to MTs mainly in tetrahedral tetrathiolate coordination spheres. From two 3D structures we know that also one or more Cys residues can be replaced by His residues without alterations of the coordination number or geometry [13, 14]. Two basic metal-thiolate structures are known: The $M^{II}_4Cys_{11}$ cluster of the α-domain of vertebrate and echinoderm MTs [15-23] and the $M^{II}_3Cys_9$ cluster of the β-domain of vertebrate, crustacean, and echinoderm MTs (Fig. 2) [15-17, 20-25].

>> insert Fig. 2 here (single column) <<

The latter cluster type is also part of the β_E-domain of wheat E_c-1 [14]. Recently, a $M^{II}_2Cys_6$ cluster was characterized by NMR spectroscopy, which was previously unprecedented for the MT superfamily and is hitherto uniquely found in the plant E_c proteins [26]. Based on the metal ion-to-Cys ratios observed in these cluster structures, i.e., 1:2.75 and 1:3, and assuming that also in plant MTs tetrahedral tetrathiolate coordination spheres prevail, expected Zn^{II} and

Cd^{II} contents can be estimated according to the number of Cys residues in the sequence as given in Table 1.

>> insert Table 1 here (single column) <<

The same rationale was followed to assess the Cu^{I} content of plant MTs. From vertebrate MTs three Cu^{I}-to-Cys stoichiometries are known, a Cu_6Cys_{11} species found in the α-domain as well as a Cu_4Cys_9 and a Cu_6Cys_9 form in the β-domain [27, 28]. No structural information shining light on the possible coordination geometries of the Cu^{I} ions in theses arrangements are available. The X-ray structure of a truncated form of the Cu^{I}-specific CUPI MT from the yeast *Saccharomyces cerevisiae* reveals a Cu_8Cys_{10} cluster, in which six Cu^{I} ions are coordinated by three Cys thiolate groups each and two show a linear coordination environment [29]. Based on these cluster forms the Cu^{I}-to-Cys ratios given in Table 1 were calculated. Both, the metal ion-to-Cys ratios for the monovalent and divalent metal ions are provided as a guideline to estimate if the metal ion contents listed in the literature for the different plant MTs can be brought into agreement with the cluster structures known from other MT species.

As the majority of studies on plant MTs are restricted to the nucleic acid level, e.g. on gene expression studies, investigations dealing with the actual proteins are rather limited in the literature. Except for the seed-specific wheat E_c-1 protein, sufficient amounts of plant MTs for more detailed analyses were only obtained via overexpression in *E. coli*. Mostly due to difficulties associated with the detection of MTs in the cell extract and to proteolytic digestion, plant MTs are largely overexpressed in form of fusion proteins, either with a GST or a self-cleavable Intein tag. As cleavage of the GST tag requires proteases, e.g. thrombin, and because of the simplicity of handling the entire fusion protein, often plant MTs are investigated in their GST-fusion form (see Table 1). From the plant MT1 subfamily the proteins from *Cicer arietinum* (chickpea), *Pisum sativum* (garden pea), and *Triticum durum* (durum wheat) have been analysed in sufficient detail to enable the determination of metal ion

contents [5, 30-33]. As summarized in Table 1, Zn^{II}, Cd^{II}, and Cu^{I} contents are mostly within the predicted range, only the values around six for the amount of coordinated divalent metal ions in garden pea are slightly higher than what would be expected based on the known cluster structures, and the Cu^{I} content of 2.3 in one preparation of garden pea MT1 is considerably lower than calculated. For the three different plant MT2 forms studied in this respect so far, the picture is even more heterogeneous. While values observed for chickpea MT2 fall into the predicted range [34, 35], metal ion contents measured in samples of *Quercus suber* (cork oak) MT2 are slightly lower (Zn^{II} and Cu^{I}) or higher (Cd^{II}) than estimated [7], and the numbers obtained for the GST-fusion protein of *Avicennia marina* (grey mangrove) are persistently below the expected range [36]. Cork oak MT2 contains an additional, not conserved His residue in the Cys-devoid linker region that was suggested to take part in Cd^{II} but only negligible in Zn^{II} binding [37]. With this His residue the number of potential Cd^{II} binding ligands would increase to 15 and hence the calculated divalent metal ion content would be 5.0–5.5. However, the Cd^{II} content of cork oak MT2 would be still higher than estimated. For the plant MT3 subfamily only the Zn^{II} and Cd^{II} contents from two forms are available: *Musa acuminata* (banana) MT3 and *Elaeis guineensis* (oil palm) MT3A [6, 38]. The calculated range based on 10 Cys residues is 3.3–3.6 metal ions. Sequence alignments show that in some plant MT3 forms, among them banana MT3 and oil palm MT3A, an additional His residue in close proximity to the C-terminal Cys-rich region is present. Supposing that this His residue could potentially serve as an additional ligand similar to the situation observed in cork oak MT2 the calculated divalent metal ion content for these MT3 forms would even increase to 4. Accordingly, coordination of three or four divalent metal ions as observed for banana MT3 is well within the predicted range. In contrast, the 1.7 Zn^{II} ions observed for oil palm MT3A are rather low. The E_c proteins constitute the plant MT subfamily with the highest number of Cys residues. As we know from the NMR solution structures of wheat E_c-1, also the two fully conserved His residues participate in metal ion

binding. Hence considering 19 potential coordinating ligands, the predicted metal ion content range in Table 1 would even increase to 6.3–6.9 divalent metal ions. However, among others the NMR structures clearly show that wheat E_c-1 has the ability to coordinate six Zn^{II} ions, and that the metal ion content is slightly lower than predicted due to the presence of a $ZnCys_2His_2$ site, i.e., a site with a metal ion-to-ligand ratio of 1:4. In light of this, a content of 2.4 Zn^{II} ions as determined for *Sesamum indicum* (sesame) clearly falls out of the range [39].

In summary, simply on the basis of the different metal ion contents observed for a given subfamily it is more than obvious that there is still a long way to go to understand the metal ion binding properties of plant MTs. Certainly it is premature to judge the accuracy of the different measurements based on the predicted values. As shown, metal ion binding capacities can be increased by recruiting additional ligands, e.g. His but also the carboxy groups of Asp and Glu would be feasible, or they can be reduced upon formation of alternative cluster structures or even mononuclear binding sites. Table 1 however, can be used as a guide to identify less likely and mostly too low metal ion contents, and it is more than obvious that there is the need for additional studies to obtain a more uniform, and hence reliable, picture for each subfamily.

Secondary structural elements

Apart from wheat E_c-1 [14, 26], no three-dimensional structural information about plant MTs is available. This means in particular that the structure and folding of the exceedingly long Cys-free linker regions so typical for the plant MT1, MT2, and MT3 proteins remain enigmatic. However, first conclusions can be drawn from results obtained with CD, IR, and Raman spectroscopy and also based on theoretical predictions and calculations, if evaluated critically. To obtain a first informative basis, the amino acid sequences of representative

members of the four plant MT subfamilies were analysed with the PROF prediction tool of the PredictProtein server (http://www.predictprotein.org) [40, 41]. When evaluating the data at the 50% probability level no helical content is predicted in any of the proteins tested, but the contribution of β-sheets is clearly evident and mostly restricted to the areas of the linker regions. Confining the β-sheet structures to the Cys-free linker regions, which account for ~50–55% of the total number of amino acids in the MT1, MT2, and MT3 forms, overall β-sheet contributions of roughly 30% are obtained (Table 2). Considering a 40% probability level these values increase to around 40%. Formation of β-sheet structures by residues of the Cys-rich regions is unlikely due to the steric strain imposed on the protein backbone by the metal-thiolate clusters. For wheat E_c-1 no α-helices and only very low β-sheet contents (below 5%) were predicted, which goes along with results of the structural investigations by NMR spectroscopy of the individual domains.

>> insert Table 2 here (single column) <<

Two studies that address the question of secondary structural elements in plant MTs with the aid of CD, IR, or Raman spectroscopy are available, the investigation of cork oak MT2 and chickpea MT1 [32, 42]. The contents of secondary structural elements in the Zn^{II} and Cd^{II} forms of cork oak MT2 were determined with Gaussian fitting of the amide I band in the Raman spectra and the amide III band in the IR spectra. While the virtual absence of helical contributions was confirmed, relatively high amounts of β-sheets (55–64%) were calculated (Table 2). A similar picture was obtained for chickpea Cd_5MT1 upon evaluation of the amide I band in the IR spectra, although the contribution of β-sheets was lower (~30%). The CD spectrum of the same species is dominated by a single minimum around 200 nm indicative for a predominantly random coil structure. Calculations revealed additionally around 10% α-helix and 30% β-sheet contributions. Taken together, the results obtained on cork oak MT2 and chickpea MT1 with the different techniques do not seem to correlate well. Hence it is

important to consider the specific weaknesses of all three spectroscopic techniques especially with respect to the study of MTs before making conclusive predictions on the amount of secondary structural elements present. The CD spectral features for α-helices and β-sheets are influenced and at least partially neutralized by the contributions of the metal-thiolate clusters, i.e. the Cd^{II}-thiolate cluster of chickpea Cd_5MT1 in the above example, tainting the calculated fractions of secondary structural elements with a relatively high uncertainty. The interpretation of the amide I and III bands in the IR and Raman spectra can be biased significantly by the parameters chosen to treat the data, e.g. smoothing and differentiation. Also the fitting procedure to decipher the individual components of the observed band shape strongly depends on the parameters applied, e.g. number of Gaussian peaks used to fit the data, their band widths and positions. Another source of ambiguity poses the assignment of the fitted Gaussian peaks to individual secondary structural elements. Overlapping spectral contributions to the amide I band in the Raman spectra have for example been described for β-sheet and β-turn structures [43]. Specifically for the amide I band in IR spectra of mammalian MTs it was shown that contributions usually assigned to α-helical structures, i.e. 1648-1660 cm^{-1}, can be also interpreted as random coil, and the wavenumber range typical for antiparallel β-sheets or aggregated strands, i.e. 1675-1695 cm^{-1}, can be occupied by contributions from β-turns in MTs [44-46].

Nevertheless and whatever the exact percentages of secondary structural elements are, it is apparent that in the plant MT1, MT2, and MT3 subfamilies virtually no α-helices are present, roughly one third of the respective entire amino acid chain consists of β-sheets and the remainder is made up of a mixture of β-turns and random coil. While the latter two structures are presumably predominantly found in the Cys rich regions, β-sheets are more conceivable in the linker regions, what is corroborated by the predictions mentioned above. It might be worth to speculate about possible arrangements of these β-sheet structures in the linker regions. An

arrangement with two longer antiparallel β-strands as depicted in Fig. 3a and predicted for *Actinidia deliciosa* (kiwi) MT3 [47] would be rather rigid and should keep the N- and C-terminal Cys-rich regions in close proximity.

>> insert Fig. 3 here (single column) <<

A relatively high β-sheet content could be alternatively achieved by multiple shorter β-sheet stretches as similarly found in the 15 amino acids long linker region of a bacterial MT [13]. This would provide the linker region with a greater flexibility potentially allowing both, a single metal-thiolate cluster (Fig. 3b) as well as two clusters formed separately by the N- and C-terminal Cys rich regions (Fig. 3c). While a single cluster has been predicted e.g. for cork oak MT2 [48] and chickpea MT2 [35], dynamic structures were proposed for e.g. chickpea MT1 [32] and banana MT3 [38], in which the formation of a single or two separate clusters is dependent on the metal ion load (see below and Fig. 4)

Influence and potential role of the Cys-devoid linker region

Linker regions in multidomain proteins can have a dual role. As a protein domain is usually considered to constitute an independent folding unit, inter-domain linker can serve as relatively rigid spacers to allow correct domain folding by preventing disturbing interferences with neighboring domains. Nevertheless, inter-domain interactions or formation of specific clefts for e.g. substrate binding can be crucial for proper protein activity. Hence, a linker can also have the function of a hinge allowing specific movements. Statistical analyses showed that Pro and Lys residues are more frequent in inter-domain linker regions [49]. In particular the occurrence of Pro demands attention, as this amino acid on the one hand interrupts the formation of secondary structural elements, but on the other hand itself introduces conformational rigidity and could assist in orientating two domains relative to each other for proper function [50]. Statistics also revealed that inter-domain linker mostly contain an even

number of amino acids ranging from 4-12 residues in length [49].

In mammalian MTs the two domains are separated by a short and highly conserved LysLysSer linker. As the linker is not involved in interdomain contacts it was suggested that it serves as a hinge region allowing a certain degree of domain movement [51]. The positive charge of the linker might also be required for charge neutralization as both domains bear a negative charge in their fully metalated form, i.e. –2 for the α- and –3 for the β-domain when coordinated to four or three divalent metal ions, respectively. To study the effect of the length of the linker region, additional amino acid repeats consisting of Pro, Gly, and Asp residues were inserted C-terminal of the LysLysSer linker and metal tolerance assays performed in form of yeast complementation assays with accordingly transformed cell lines [51]. Insertion of up to 4 additional amino acids confers the same tolerance against up to 300 μM Cd^{II} ions as observed for the unmodified mammalian MT, constructs with longer inserts, i.e. 8, 12, and 16 amino acids, were less effective. As concomitantly also lower levels of the respective MT proteins with the longer linkers were observed, a reduced half-life of these proteins in yeast was proposed and hence a higher propensity of the longer linkers to proteolytic degradation. Thus from this study it is not possible to decipher if the increased spatial separation of the two domains in the investigated mammalian MT per se decreases the metal ion binding ability of the protein or if only the actual proteolytic cleavage of the linker during protein degradation diminishes the activity in the metal tolerance assay. The second point not addressed is the influence of the charge of the linker residues on the metal ion binding properties. Firstly, the insertion of additional linker residues C-terminal of the LysLysSer sequence increases the distance of the α-domain to the +2 charge of the linker and secondly, the longer inserts of 8, 12, and 16 amino acids contain 2, 3, and 4 Asp residues, respectively, adding to the overall negative charge. On a speculative basis, these additional negative charges might destabilize the metal-thiolate clusters – independently of the linker length. Hence the observed reduced metal ion binding properties of mammalian MT upon increase of the linker length clearly

requires further investigation before justifying a general conclusion. Differences in properties however are apparent when comparing the full-length human MT2 protein with its two separate domains *in vitro*. The fully metal ion loaded full-length protein was found to be more stable towards oxidation and Zn^{II} transfer [52], and remetallation studies with the apo-forms and As^{III} revealed the faster metallation of the full-length form [53]. In addition, the apparent pK_a values of the Cys residues in the Zn^{II}- and the Cd^{II}-forms, which are indicative for the metal ion affinities of the proteins, are lower for the full-length human MT2 compared to the average values obtained from separate measurements with the individual domains [52]. In summary, while the influence of the length of the linker region is less clear as variations of the amino acid composition might have a difficult to predict influence, it seems obvious that the physical connection of both domains plays a crucial role for the stability of the protein.

Turning our attention to the plant MTs, one of the most striking features are their unusual long Cys-devoid linker regions ranging from ~30-45 amino acids in length for the MT1, MT2, and MT3 forms, while the linker between the γ- and β_E-domains of the E_c proteins is with up to 11 residues – when considering the conserved His residues as part of the central Cys-rich region – considerably shorter. Up to now, solely a single study can be found in the literature providing experimental results on the role of the Cys-free linker regions in plant MTs [48]. The comparative study showed that both full-length cork oak MT2 as well as a chimera of the N- and C-terminal Cys-rich regions connected by a shorter 8 amino acids long linker are able to form species with the same amount of metal ions, i.e. Zn_4- and Cu_8-forms. Interestingly and despite of the identical metal ion binding capacity, the full-length protein is superior to the species with the shorter linker in yeast complementation studies for copper tolerance. Reasons for this might be an influence of the longer linker on the *in vivo* stability of the full-length protein, a role in targeting of the protein to a specific subcellular location or even interaction with a so-far unknown binding partner within the cell [48]. Contrarily, earlier studies raised the possibility that plant MTs are posttranslationally processed to remove the linker region

based on trials to isolate the native proteins directly from the plant material [54]. Initial failure to overexpress garden pea MT1 in *E. coli* supported this view [5]. In analogy to results obtained with the full-length human MT2 and its separate domains, also the apparent pK_a values of the Cys residues in full-length wheat E_c-1 are lower than the average of the values obtained for the separately analyzed γ- and β_E-domains of the protein [26]. This averaged value is identical to the pK_a values obtained, when a 1:1 mixture of the separate domains was subjected to pH titration. Again it seems to be the actual physical connection of the two domains by the linker region that is crucial for the increased metal ion binding stabilities of the full-length protein.

Spatial arrangement of Cys-rich regions

After discussing the secondary structural elements presumably formed by the residues of the linker regions and the importance of these linkers for the properties and functions of (plant) MTs, we will now focus on the spatial arrangement of the Cys-rich regions and hence on the metal-thiolate cluster assemblies formed. Beyond doubt, the most intensely investigated and best understood structure is the one of the mammalian MT forms. The N-terminal β- and the C-terminal α-domain can be regarded as independent folding units in the presence of suitable metal ions, hence the final structure is identical in the separately studied domains and the full-length protein [15]. The domains are connected by a LysLysSer sequence as mentioned above, whose linker nature became only clear in the course of the structural studies. Consequently, mammalian MTs are two-domain proteins in the widest sense and their structure is commonly described as dumbbell-shaped. The plant MT1, MT2, and MT3 proteins analogously feature two Cys-rich regions, and their separation in sequence by the long linker regions is obvious (Fig. 1). However, no three-dimensional information about the structures of members from these three plant MT subfamilies is available, and accordingly the question concerning the

spatial arrangement of the two Cys-rich regions relative to one another has to rely on alternative techniques. In principle both, a dumbbell-shaped two-domain structure as in the mammalian forms or a hairpin-like one-domain arrangement as seen for example in a cyanobacterial MT [13] is feasible. The earliest study providing insights into this question describes the overexpression of garden pea MT1 without purification tag in *E. coli*, which was accompanied by proteolytic cleavage of the linker region as mentioned above [5]. Accordingly, reverse phase chromatography at pH 2 revealed two distinct peaks originating from the N- and C-terminal Cys-rich regions, respectively. However, only a single peak was observed after size exclusion chromatography at neutral pH and in presence of Cd^{II} ions. This triggered the obvious conclusion that both peptides were connected by Cd^{II} ions even after cleavage of the protein backbone and hence a single metal ion cluster must be present in the protein. In a similar way, the spatial arrangement of the two Cys-rich regions was addressed for chickpea MT2 [35]. Limited proteolytic digestion of the Cd_5MT2 form with proteinase K and subsequent size exclusion chromatography gave a single peak with an elution time corresponding to an apparent molecular mass of around 4.5 kDa, hence the mass of the combined Cys-rich regions. MALDI-TOF measurements of the peak fraction identified solely signals for the two Cys-rich domains, but not for the linker region. Amino acid analysis corroborated the result. Hence it is save to conclude that also chickpea MT2 adopts a hairpin-like structure in its fully metalated form. Also for cork oak MT2 a hairpin-like structure with a single metal ion cluster was proposed based on a comparative analysis of CD spectra [48].

In contrast to the results above is the theoretical work performed with the sequence of durum wheat MT1 [30]. The authors used both Cys-rich regions, which contain six Cys residues each, separately for homology modelling yielding the β-domains of sea urchin MT and rat liver MT2, respectively, as the best fitting models. Both models contain a Cd_3Cys_9 cluster and hence prompted the prediction of a dumbbell-shaped arrangement of the two Cys-rich regions of durum wheat MT1 with two separate Cd_3Cys_6 clusters. It has to be noted however that this

result (i) contrasts the metal ion content of durum wheat MT1 that was determined to 4 ± 1 Cd^{II} ions in the same study and is well in line with results from other plant MT1 proteins and theoretical calculations (Table 1), (ii) is hard to reconcile with any metal-thiolate cluster stoichiometries and structures known so far, and (iii) is not in line with the hairpin-like spatial arrangement predicted for garden pea MT1. The proposal of different cluster arrangements for the two MT1 proteins is even more surprising as durum wheat MT1 shows 52 % sequence identity and 73 % similarity to the sequence of garden pea MT1. Considering only the Cys-rich regions the sequence similarity is even 100 %.

Banana MT3 has the ability to coordinate up to four divalent metal ions [38]. Based on stoichiometric considerations formation of a hairpin-like structure seems reasonable as banana MT3 contains 10 Cys and one His residue, which in principle allows a metal cluster arrangement similar to the α-domain of the mammalian MTs ($M^{II}_4Cys_{11}$) or the cluster found in the cyanobacterial form ($M^{II}_4Cys_9His_2$). For both, banana MT3 and its His→Ala mutant, also a species coordinating only three Zn^{II} ions was observed and triggered the proposal of an additional substoichiometrically metal-loaded form that could adopt a dumbbell-shaped structure [38]. In this model, the N-terminal region would bind a single Zn^{II} ion via its four Cys residues, while a separate Zn_2Cys_6 cluster could be accommodated in the C-terminal Cys-rich region (Fig. 4a).

>> insert Fig. 4 here (single column) <<

This view opens up the possibility of a jack-knife like domain movement in dependence of the metal ion load. Coordination of the forth metal ion would then accordingly bring the two separate clusters together to form a single metal cluster arrangement (Fig. 4b). A two-domain, "open" structure was similarly predicted for kiwifruit MT3 in its three divalent metal ions binding form based on theoretical calculations [47].

The fourth plant MT subfamily, the E_c proteins, contain even three Cys-rich regions, again clearly separated in sequence by, although shorter, linker stretches. Corroborated by a number

of investigations and last not least two NMR solution structures, we have solid evidence that wheat E_c-1 is organized into two metal-binding domains [14, 26, 55, 56]. The smaller N-terminal domain is formed by the N-terminal Cys-rich region and hosts a Zn_2Cys_6 cluster (Fig. 5).

>> insert Fig. 5 here <<

As mentioned above such a cluster, while known from yeast transcription factors, e.g. GAL4 [57], was unprecedented for any MT species so far. In continuation of the nomenclature used for the two mammalian MT domains, the N-terminal domain of wheat E_c-1 was denominated as γ-domain. The larger domain is formed by both the central and the C-terminal Cys-rich region and encloses a mononuclear $ZnCys_2His_2$ site as well as a Zn_3Cys_9 cluster with similarity to the one found in the β-domain of, e.g., the mammalian MTs (Fig. 5). This arrangement of metal ion binding sites was denominated as extended β-domain or β_E-domain. While in principle coordination of Zn^{II} ions by His residues is known from cyanobacterial MTs [13], the observed mononuclear binding site is again unprecedented for any MT form so far. The Cys residues of the central Cys-rich region participate both in the formation of the mononuclear site and the Zn_3Cys_9 cluster giving rise to an interleaved arrangement of the two metal ion binding sites in this domain. Both, the γ- and the β_E-domain, have been shown to act as independent folding units [26, 55], and from the NMR studies on the full-length protein no interdomain contacts could be deduced [14]. Hence, the overall arrangement of the wheat E_c-1 protein might be described as elongated or dumbbell-shaped. Taken together, it is more than legitimate to say that such an arrangement of the three Cys-rich regions would have been more than difficult to predict with theoretical methods, especially given the novelty of the structural motifs. Nevertheless, the basic motif of a two-domain protein formed by the N-terminal Cys-rich region and the combined central and C-terminal Cys-rich regions was already predicted based on proteolytic digestion experiments as the main method [56] and also contemplated while studying pH-dependent Zn^{II} release from the full-length protein with

ESI-MS [58].

This review of the literature shows that prediction of the spatial arrangement of Cys-rich regions in any new MT species based on theoretical considerations or homology modelling bears a certain degree of ambiguity. On the other hand the relatively simple experiment of limited proteolytic digestion coupled with appropriate analysis of the resulting cleavage product(s) can give an ample first insight into the prevailing cluster organisation. This being said, the actual metal-thiolate cluster structure, and possibly the participation of other coordinating ligands than Cys residues, might still hold some surprises at hand.

Acknowledgements This work was supported by the Swiss National Science Foundation (SNSF Professorship PP002-119106/1 to E.F.).

References

1. Margoshes M, Vallee BL (1957) J Am Chem Soc 79:4813-4814
2. Hanley-Bowdoin L, Lane BG (1983) Eur J Biochem 135:9-15
3. Lane BG, Kajioka R, Kennedy TD (1987) Biochem Cell Biol 65:1001-1005
4. Kawashima I, Kennedy TD, Chino M, Lane BG (1992) Eur J Biochem 209:971-976
5. Kille P, Winge DR, Harwood JL, Kay J (1991) FEBS Lett 295:171-175
6. Abdullah SNA, Cheah SC, Murphy DJ (2002) Plant Physiol Bioch 40:255-263
7. Mir G, Domènech J, Huguet G, Guo W-J, Goldsbrough P, Atrian S, Molinas M (2004) J Exp Bot 55:2483-2493
8. Kägi JHR, Himmelhoch SR, Whanger PD, Bethune JL, Vallee BL (1974) J Biol Chem 249:3537-3542
9. Kojima Y, Berger C, Vallee BL, Kägi JHR (1976) Proc Natl Acad Sci U S A 73:3413-3417

10. Kojima Y, Binz P-A, Kägi JHR (1999) In: Klaassen C (ed) Metallothionein IV. Birkhäuser Verlag, Basel, pp. 3-6
11. Freisinger E (2008) Dalton Trans:6663-6675
12. Freisinger E (2009) Met Ions Life Sci 5:107-153
13. Blindauer CA, Harrison MD, Parkinson JA, Robinson AK, Cavet JS, Robinson NJ, Sadler PJ (2001) Proc Natl Acad Sci U S A 98:9593-9598
14. Peroza EA, Schmucki R, Güntert P, Freisinger E, Zerbe O (2009) J Mol Biol 387:207-218
15. Braun W, Vašák M, Robbins AH, Stout CD, Wagner G, Kägi JHR, Wüthrich K (1992) Proc Natl Acad Sci U S A 89:10124-10128
16. Riek R, Prêcheur B, Wang Y, Mackay EA, Wider G, Güntert P, Liu A, Kägi JHR, Wüthrich K (1999) J Mol Biol 291:417-428
17. Capasso C, Carginale V, Crescenzi O, Di Maro D, Parisi E, Spadaccini R, Temussi PA (2003) Structure (Camb) 11:435-443
18. Öz G, Zangger K, Armitage IM (2001) Biochemistry-Us 40:11433-11441
19. Wang H, Zhang Q, Cai B, Li HY, Sze KH, Huang ZX, Wu HM, Sun HZ (2006) FEBS Lett 580:795-800
20. Arseniev A, Schultze P, Worgotter E, Braun W, Wagner G, Vašák M, Kägi JHR, Wüthrich K (1988) J Mol Biol 201:637-657
21. Messerle BA, Schaffer A, Vašák M, Kägi JHR, Wüthrich K (1990) J Mol Biol 214:765-779
22. Schultze P, Worgotter E, Braun W, Wagner G, Vašák M, Kägi JHR, Wüthrich K (1988) J Mol Biol 203:251-268
23. Zangger K, Öz G, Otvos JD, Armitage IM (1999) Protein Sci 8:2630-2638
24. Muñoz A, Försterling FH, Shaw III CF, Petering DH (2002) J Biol Inorg Chem 7:713-724

25. Narula SS, Brouwer M, Hua Y, Armitage IM (1995) Biochemistry-Us 34:620-631
26. Loebus J, Peroza EA, Blüthgen N, Fox T, Meyer-Klaucke W, Zerbe O, Freisinger E (2011) J Biol Inorg Chem 16:683-694
27. Faller P, Vašák M (1997) Biochemistry-Us 36:13341-13348
28. Nielson KB, Winge DR (1984) J Biol Chem 259:4941-4946
29. Calderone V, Dolderer B, Hartmann HJ, Echner H, Luchinat C, Del Bianco C, Mangani S, Weser U (2005) Proc Natl Acad Sci U S A 102:51-56
30. Bilecen K, Ozturk UH, Duru AD, Sutlu T, Petoukhov MV, Svergun DI, Koch MHJ, Sezerman UO, Cakmak I, Sayers Z (2005) J Biol Chem 280:13701-13711
31. Schicht O:PhD Thesis, University of Zurich, 2007
32. Schicht O, Freisinger E (2009) Inorg Chim Acta 362:714-724
33. Tommey AM, Shi J, Lindsay WP, Urwin PE, Robinson NJ (1991) FEBS Lett 292:48-52
34. Wan X:PhD Thesis, University of Zurich, 2009
35. Wan X, Freisinger E (2009) Metallomics 1:489-500
36. Huang G-Y, Wang Y-S (2010) Aquat Toxicol 99:86-92
37. Torreggiani A, Domènech J, Tinti A (2009) J Raman Spectrosc 40:1687-1693
38. Freisinger E (2007) Inorg Chim Acta 360:369-380
39. Chyan CL, Lee TT, Liu CP, Yang YC, Tzen JT, Chou WM (2005) Biosci Biotechnol Biochem 69:2319-2325
40. Rost B, Fariselli P, Casadio R (1996) Protein Sci 5:1704-1718
41. Rost B, Sander C (1993) J Mol Biol 232:584-599
42. Domènech J, Tinti A, Capdevila M, Atrian S, Torreggiani A (2007) Biopolymers 86:240-248
43. Alix AJP, Pedanou G, Berjot M (1988) J Mol Struct 174:159-164
44. Jackson M, Mantsch HH (1995) Crit Rev Biochem Mol 30:95-120

45. Byler DM, Susi H (1986) Biopolymers 25:469-487

46. Shi YB, Fang JL, Liu XY, Du L, Tang WX (2002) Biopolymers 65:81-88

47. Zhu C, Lü T, Zhang R, Zhao N, Liu J (2000) Chinese Sci Bull 45:1413-1417

48. Domènech J, Mir G, Huguet G, Capdevila M, Molinas M, Atrian S (2006) Biochimie 88:583-593

49. Tanaka T, Kuroda Y, Yokoyama S (2003) J Struct Funct Genomics 4:79-85

50. Miyazaki S, Kuroda Y, Yokoyama S (2002) J Struct Funct Genomics 2:37-51

51. Rhee IK, Lee KS, Huang PC (1990) Protein Eng 3:205-213

52. Jiang LJ, Vašák M, Vallee BL, Maret W (2000) Proc Natl Acad Sci U S A 97:2503-2508

53. Ngu TT, Easton A, Stillman MJ (2008) J Am Chem Soc 130:17016-17028

54. de Miranda JR, Thomas MA, Thurman DA, Tomsett AB (1990) FEBS Lett 260:277-280

55. Peroza EA, Al Kaabi A, Meyer-Klaucke W, Wellenreuther G, Freisinger E (2009) J Inorg Biochem 103:342-353

56. Peroza EA, Freisinger E (2007) J Biol Inorg Chem 12:377-391

57. Baleja JD, Thanabal V, Wagner G (1997) J Biomol NMR 10:397-401

58. Leszczyszyn OI, Schmid R, Blindauer CA (2007) Proteins: Structure, Function, and Bioinformatics 68:922-935

59. dos Santos Cabral A:PhD Thesis, University of Zurich, 2010

Table 1 Metal ion contents of plant MTs as calculated based on metal ion-to-protein ratios found in the metal-thiolate clusters of vertebrate MTs and as determined experimentally on the purified proteins.[a]

Species		Zn^{II}-form	Cd^{II}-form	Cu^{I}-form	Ref.
MT1 (12 Cys)[b]		4.0–4.4	4.0–4.4	5.3–8.0 (9.6)	
C. arietinum (chickpea)	MT1	4, 5	4, 5	6, 9	[31, 32]
P. sativum (garden pea)	MT1	5.6[c]	3.9[c], 5.8[c]	2.3[c], 6.2[c]	[33]
			5.6–6.1		[5]
T. durum (durum wheat)	MT1		4 (±1)[c]		[30]
MT2 (14 Cys)[b]		4.7–5.1	4.7–5.1	6.2–9.3 (11.2)	
C. arietinum (chickpea)	MT2	5	5	8	[34, 35]
Q. suber (cork oak)	MT2	3.5	6.3	5.5 (+1.7 Zn^{II})	[7]
A. marina (grey mangrove)	MT2	3[c]	3[c]	4[c]	[36]
MT3 (10 Cys)[b]		3.3–3.6	3.3–3.6	4.4–6.7 (8.0)	
M. acuminata (banana)	MT3	3, 4	4		[38]
E. guineensis (oil palm)	MT3-A	1.7[c]			[6]
E_c (17 Cys)[b]		5.7–6.2	5.7–6.2	7.6–11.3 (13.6)	
T. aestivum (bread wheat)	E_c-1	5			[3]
		6	6		[56],[58][d]
S. indicum (sesame)	E_c	2.4 (±0.6)		2.2 (±0.5)	[39]

[a] Adapted from [11] - Reproduced by permission of The Royal Society of Chemistry (RSC). [b] Range calculated based on the metal ion-to-cysteine ratios in the α-domain ($M^{II}_4Cys_{11}$, 1:2.75; $M^{I}_6Cys_{11}$, 1:1.83) and β-domain ($M^{II}_3Cys_9$, 1:3; $M^{I}_6Cys_9$, 1:1.5; $M^{I}_4Cys_9$, 1:2.25) of vertebrate MTs; for the Cu^{I}-content in addition also the $Cu^{I}_8Cys_{10}$ cluster of yeast CUPI was considered (1:1.25; numbers in brackets). [c] GST-fusion protein. [d] 5.8(6) Zn^{II} and 0.6(1) $Cu^{I/II}$ ions were found in the protein directly isolated from wheat germs.

Table 2 β-sheet and α-helix contents of selected plant MTs as predicted for the linker regions or determined with IR, Raman, or NMR spectroscopy for the entire proteins. All numbers given refer to % of amino acids in the full-length proteins

		Predicted[a]		Determined (M^{II}-form)		
Species	% linker aa	β-sheet	α-helix	β-sheet	α-helix	Ref.
C. arietinum MT1	56	27/41	0/0	26-30	0	[32]
Q. suber MT2	51	27/39	0/0	55-64	0-4	[42]
M. acuminata MT3	52	29/38	0/3	31-41	0	[59]
T. aestivum E_c-1	32	2/5	0/0	0	0	[14, 26]

[a] The predicted values are listed for the 50 and 40% probability level (50/40), respectively [40, 41].

Figure legends

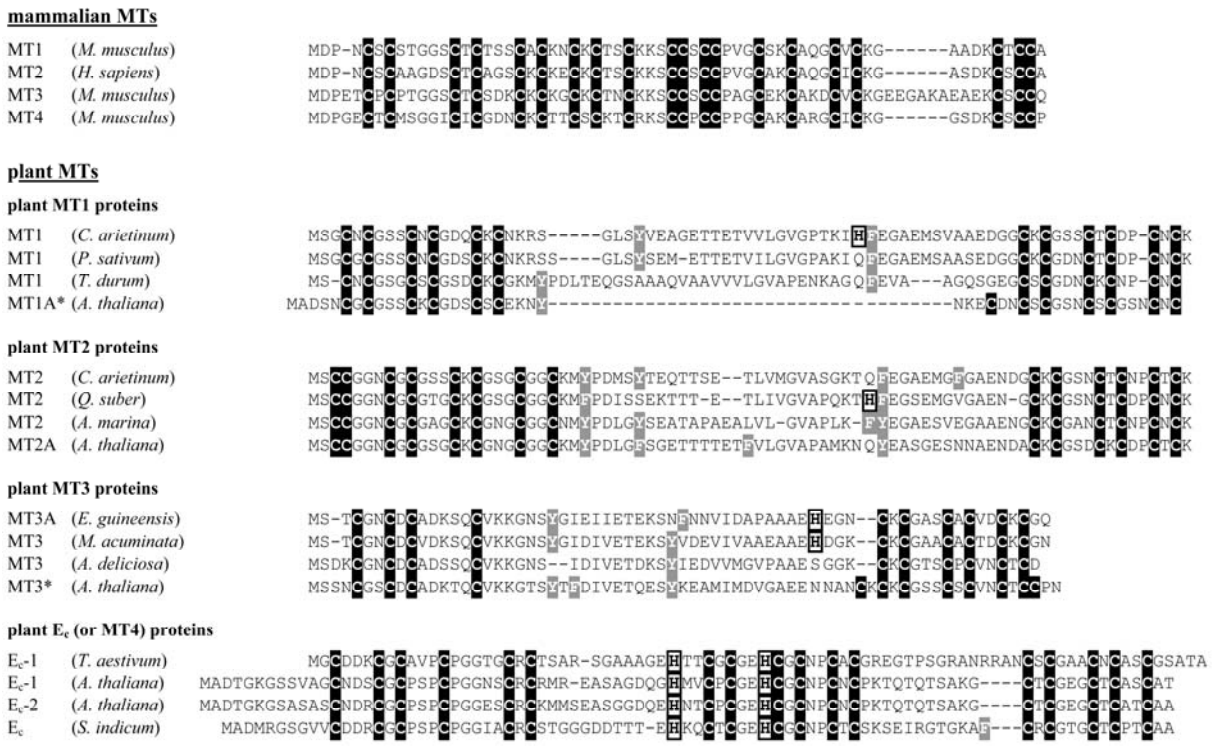

Fig. 1 Amino acid sequence alignment of representative members of the mammalian and the plant MT families. The plant MTs are additionally divided into four subfamilies comprising the MT1, MT2, MT3, and E_c proteins. Cys residues are highlighted with a black background, aromatic amino acids with a grey background. His residues are accentuated with a black frame. Sequences denoted with * represent exceptions to the otherwise highly conserved Cys distribution pattern within a plant MT subfamily. In *A. thaliana* MT1A the linker region is additionally reduced to just seven amino acids.

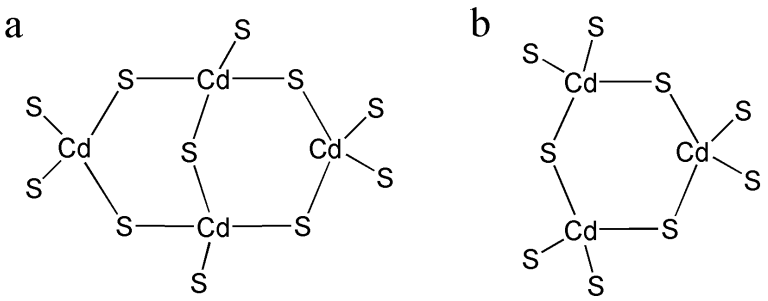

Fig. 2 Schematic presentation of the two metal-thiolate cluster structures formed with divalent metal ions in e.g. the vertebrate MTs: **a** $M^{II}_4Cys_{11}$ cluster of the α-domain and **b** $M^{II}_3Cys_9$ cluster of the β-domain. Only the coordinating sulfur atoms of the Cys residues are shown.

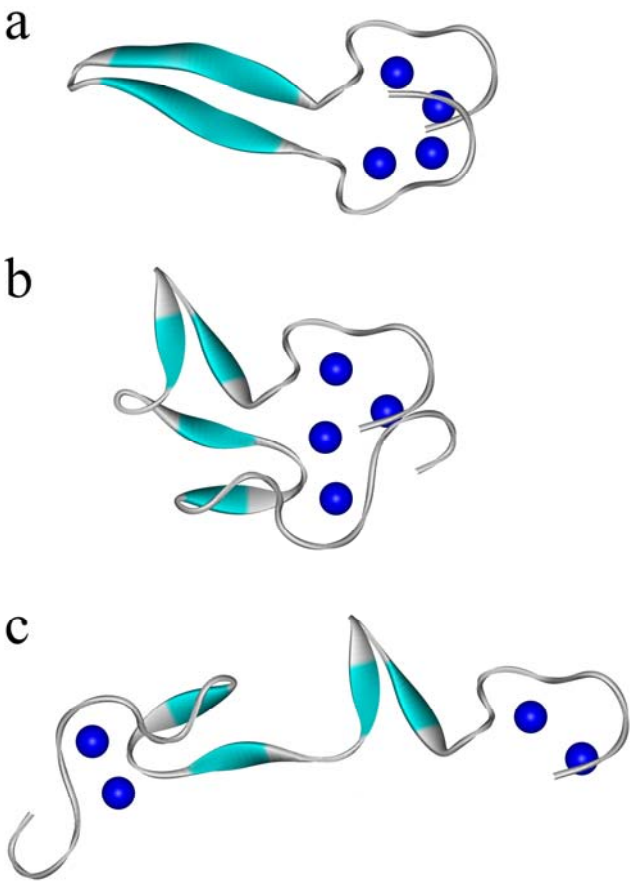

Fig. 3 Schematic presentation of probable β-sheet arrangements within the linker regions of plant MTs. **a** The entire linker sequence forms a single long antiparallel β-sheet structure (cyan) and hence a rather rigid scaffold to bring the Cys-rich regions (long, grey terminal tubes) into proximity for joint cluster formation (metal ions depicted as blue spheres). The amino acids of the linker region form a more flexible structure with four shorter β-sheets, which allow a single cluster (**b**) or a dumbbell-shaped arrangement with two clusters (**c**).

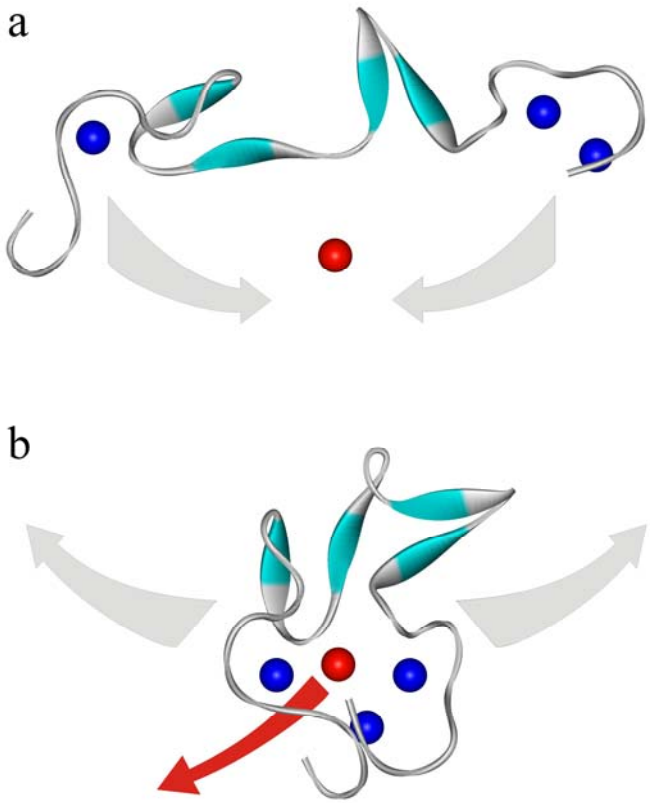

Fig. 4 a The dumbbell-shaped arrangement similar to Fig. 3a can be transformed in a jack-knife-like movement into a single domain cluster form upon binding of an additional metal ion (red sphere). **b** This transformation can be reversed upon removal of the additional metal ion. The grey arrows show the direction of movement of the Cys-rich regions, the red arrow the release of the additional metal ion.

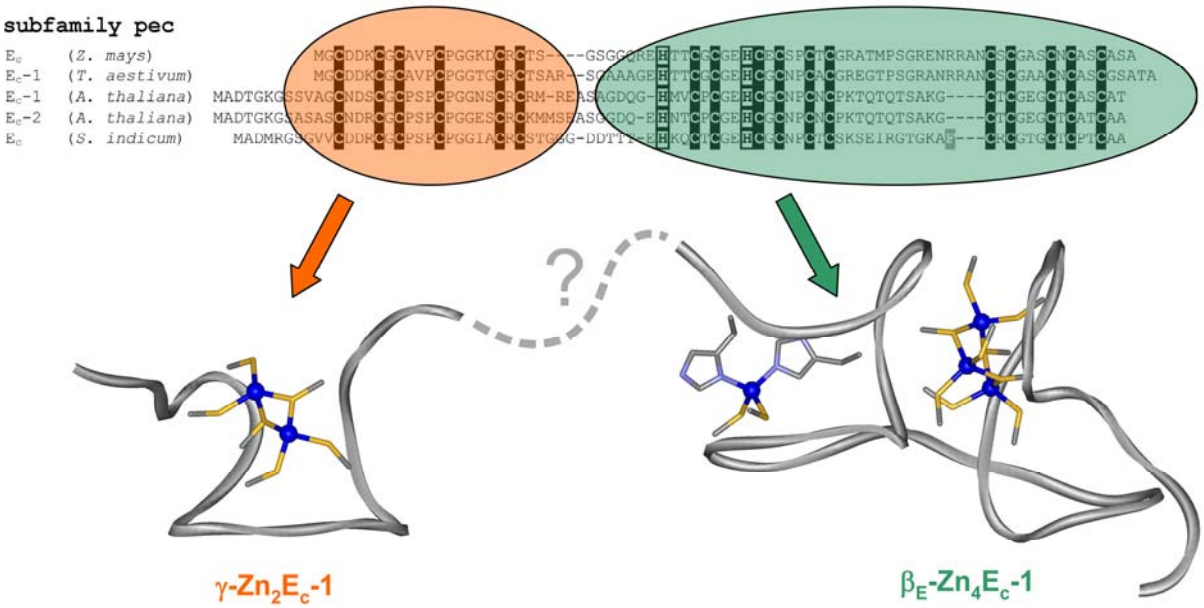

Fig. 5 Amino acid sequence alignment of representative members of the plant E_c subfamily with Cys and His residues highlighted as in Fig. 1. The residues comprising the N-terminal γ- and the C-terminal β_E-domain are indicated with an orange and green ellipsoid, respectively. Below this, the NMR solution structures of the two domains of wheat E_c-1 are shown, no information about the relative orientation of these two domains to each other is available. Zn^{II} ions are depicted as blue spheres, parts of the coordinating Cys and His side chains are shown in stick mode.